Baba Ghulam Shah Badshah University Rajouri (J&K)

(SYNOPSIS FOR REGISTRATION FOR DEGREE OF DOCTOR OF PHILOSPHY IN COMPUTER SCIENCE)

Proposed Topic	:	**"Soft Computing Based Privacy Preserving Data Mining"**
Name of the Candidate	:	Mr. Majid Bashir Malik
Name of the Supervisor	:	**Supervisor:** Prof. M. Asger Department of Computer Sciences, Baba Ghulam Shah Badshah University, Rajouri **Co-Supervior:** Dr. Rashid Ali Department of Computer Engineering, Aligarh Muslim University, Aligarh.
School/Department	:	Computer Sciences, School of Mathematical Sciences & Engineering.
Date of Submission	:	30-11-2011

Signature of the Candidate

Signature of the Supervisor

Signature of the Co-Supervisor
(Dr. Rashid Ali)

I. Background

Introduction

Rapid advances in data collection and storage technology have enabled organization to accumulate vast amounts of data. It has been estimated that the amount of information in the world doubles every 20 months and the size and number of the data bases are increasing even faster [1]. However extracting useful information has proven extremely challenging. Traditional analysis tools and techniques cannot be used because of massive size of data. Data Mining blends traditional data analysis methods with sophisticated algorithms for processing large volumes of data. The raw material is the transactional data, while, data mining algorithm is the excavator that filters the valuable nuggets of information from vast quantities of raw data [1]. Data mining is a set of automated techniques used to extract hidden or buried information from large databases. The term data mining refers to the nontrivial extraction of valid, implicit, potentially useful and ultimately understandable information in large databases with the help of the modern computing devices [2, 3]. In the last few decades, many successful applications in data mining have been reported from varied sectors such as marketing, finance, banking, manufacturing and telecommunication. Apart from the benefits of using data per se (e.g. keeping up to date profiles of the customers and their purchases, maintaining a list of available products, their quantities and prices etc.), the mining of these data sets with the existing data mining tools [4] can reveal invaluable knowledge that was unknown to the data holder before hand. The extracted knowledge patterns can provide insight to the data holders as well as be invaluable in tasks such as decision making and strategic business planning [5]. As a valuable technique, data mining is developing and is flourishing. But, at the same time, serious concerns have grown over individual privacy in data collection, processing and mining [6]. As a result, preserving privacy is a serious issue in data mining. It is perceived that data mining, with its promise to efficiently discover valuable, non-obvious information from large databases, is particularly vulnerable to misuse [7, 8]. So, there might be a conflict between data mining and privacy. But, in reality, there is no need to intrude privacy for data mining. The objective of data mining is to generalize across populations, rather than reveal information about individuals. But, the problem is that data mining works by evaluating individual data. Consequently, there might be

privacy intrusion. Hence, the true problem is not with data mining, but with the way data mining is done.

Moreover, companies are often willing to collaborate with other entities, who conduct similar business, to obtain mutual benefits of their businesses. Significant knowledge patterns can be derived and shared among the collaborative partners through aggregate mining of their datasets. Furthermore, public sector organizations and civilian federal agencies usually need to share a portion of their collected data or knowledge with other organizations having similar purpose. Sometimes, they even make this data or knowledge public. Now, it is clear that there exists an extended set of application scenarios in which data or knowledge derived from the data has to be shared with other (possibly untrustworthy) entities. The sharing of data and/or knowledge might come at the cost of privacy, primarily due to two reasons [5]:

1. if the data refers to the individual (e.g. as in patients' health data) then the disclosure of this data or any knowledge extracted from the data can potentially violate the privacy of individuals if their identity is revealed to the unreliable third parties.
2. if the data contains any business (or organizational) information, then the disclosure of this data or any knowledge extracted from data can potentially disclose sensitive trade secrets, whose knowledge can provide significant gain to business competitors and thus can cause data holder to lose business over his/her peers.

In [9] and [10], authors proposed the concept of privacy preserving data mining (PPDM) aimed at alleviating the problem between data mining and privacy. PPDM may be defined as "getting valid data mining results without learning the underlying data values" [11]. PPDM encompasses the dual goal of meeting privacy requirements and providing valid data mining results [12].

Soft computing techniques may be used to handle the different challenges offered by the data mining [13]. Soft computing is a consortium of methodologies which work synergistically and provides in one form or another flexible information processing capabilities for handling real-life ambiguous situations [14]. Methodologies of soft

computing include fuzzy logic, neural networks, genetic algorithms, and rough sets.

Fuzzy Sets

Fuzzy sets provide a natural framework for the process in dealing with uncertainty. It makes it possible to model imprecise and qualitative knowledge as well as the transmission and handling of uncertainty at various stages. It supports human type reasoning in natural form

Neural Networks

Neural Network is a highly interconnected network of a large number of processing elements called neurons in an architecture inspired by brain. Neural Networks are widely used for classification and rule generation.

Genetic Algorithms

Genetic algorithms are adaptive, robust, efficient and global search methods, suitable in situations where the search space is large. Genetic algorithms are involved in various optimization and search processes like query optimization and template selection [13].

Rough Sets

Rough set is a mathematical tool for managing uncertainty that arises from indiscernibility between objects in a set. Principally it discovers redundancies and dependencies between the given features of a problem to be classified [13].

Each of them contributes a distinct methodology for addressing problems in its domain. This is done in a cooperative, rather than a competitive, manner. It results in an intelligent and robust system providing a human-interpretable, low cost, approximate solution, as compared to traditional techniques [13]. In this work, we intend to study soft computing techniques based privacy preserving data mining.

Related Work:

As a research direction in data mining and statistical databases [15], privacy preserving data mining received substantial attention and many researchers worked in the area [9-12, 16-19, 30]. Since its inception in 2000 with the pioneering work of Agrawal & Srikant [9] and Lindell & Pinkas [10], privacy preserving data mining has gained increasing popularity in data mining research community. Privacy Preserving Data Mining (PPDM)

has become an important issue in data mining research [31-33]. As a result a whole new set of approaches were introduced to allow mining of data, while at the same time prohibiting the leakage of any private and sensitive information. The majority of the existing approaches can be classified into two broad categories [5]

1. methodologies that protect the sensitive data itself in the mining process, and
2. methodologies that protect the sensitive data mining results (i.e. extracted knowledge) that were produced by the application of the data mining.

The first category refers to the methodologies that apply perturbation, sampling, generalization/suppression, transformation, etc. techniques to the original datasets in order to generate their sanitized counterparts that can be safely disclosed to untrustworthy parties. The goal of this category of approaches is to enable the data miner to get accurate data mining results when it is not provided with the real data. Secure Multiparty Computation methodologies that have been proposed to enable a number of data holders to collectively mine their data without having to reveal their datasets to each other.

The second category deals with techniques that prohibits the disclosure sensitive knowledge patterns derived through the application of data mining algorithms as well as techniques for downgrading the effectiveness of classifiers in classification tasks, such that they do not reveal sensitive knowledge.

In contrast to the centralized model, the Distributed Data Mining (DDM) model assumes that the data sources are distributed across multiple sites. Algorithms developed within this field address the problem of efficiently getting the mining results from all the data across these distributed sources. A simple approach to data mining over multiple sources that will not share data is to run existing data mining tools at each site independently and combine the results [20, 21, 25]. However, this will often fail to give globally valid results. Issues that cause a disparity between local and global results include:

1. Values for a single entity may be split across sources. Data mining at individual sites will be unable to detect cross-site correlations.

2. The same item may be duplicated at different sites, and will be over-weighted in the results.

3. Data at a single site is likely to be from a homogeneous population. Important geographic or demographic distinctions between that population and others cannot be seen on a single site.

Cheung et al. proposed a method for horizontally partitioned data [23]. Distributed classification has also been addressed. A meta-learning approach has been developed that uses classifiers trained at different sites to develop a global classifier [20, 21, 25]. This could protect the individual entities, but it remains to be shown that the individual classifiers do not disclose private information. Recent work has addressed classification using Bayesian Networks in vertically partitioned data [22], and situations where the distribution is itself interesting with respect to what is learned [28]. Shenoy et al. [27] propose an efficient algorithm for vertically mining association rules. Data mining algorithms that partition the data into subsets have been developed [26]. Although the goal of parallelizing data mining algorithms is performance, the communication cost between nodes is an issue. Parallel data mining algorithms may also serve as a starting point [24, 29]. However, none of this work directly addresses privacy concerns. With distributed data, the way the data is distributed also plays an important role in defining the problem. Data could be partitioned into many parts either vertically or horizontally.

Vertical Partitioning

Vertical partitioning (heterogeneous distribution) of data implies that though different sites gather information about the same set of entities, they collect different feature sets. For example, financial transaction information is collected by banks, while the IRS collects tax information for everyone. An illustrative example of vertical partitioning and the kind of useful knowledge we can hope to extract is given in figure 1. The figure describes two databases, one contains medical records of people while another contains cell phone information for the same set of people. Mining the joint global database might reveal information like “Cell phones with Li/Ion batteries lead to brain tumors in diabetics.”

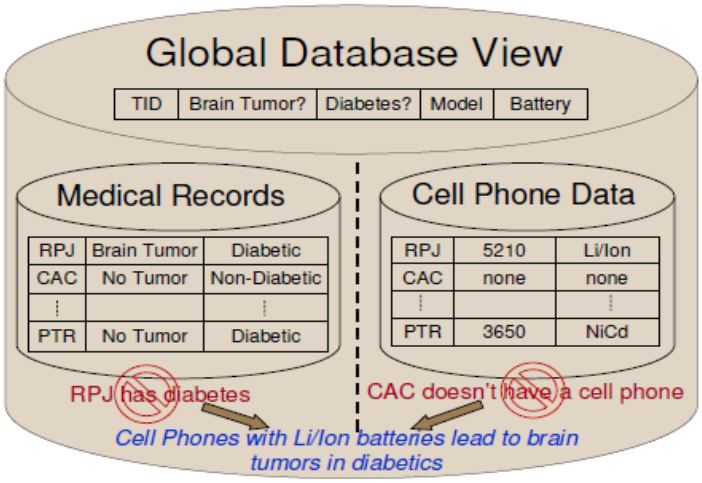

Figure 1: Vertically Partitioned Data

Horizontal Partitioning

In horizontal partitioning (homogeneous distribution), different sites collect the same set of information, but about different entities. An example of that would be grocery shopping data collected by different supermarkets (also known as market-basket data in the data mining literature). Figure 2 illustrates horizontal partitioning and shows the credit card databases of two different (local) credit unions. Taken together, one may find that fraudulent customers often have similar transaction histories, etc.

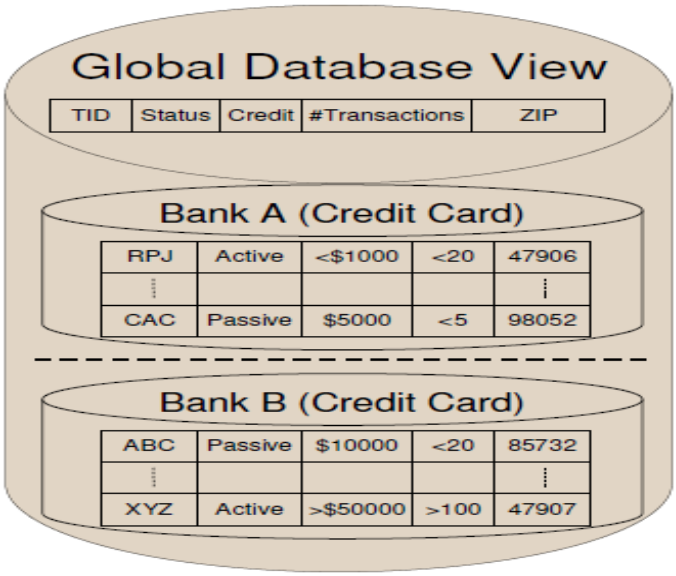

Figure: Horizontally partitioned database

These different partitioning pose different problems, leading to different algorithms for privacy-preserving data mining.

Quantification has always been an important issue in PPDM. The quantity used to measure privacy should indicate how closely the original value of an attribute can be estimated [44]. In [44] a reconstruction algorithm for PPDM has been proposed which not only converges but does so to maximum likelihood estimate of original distribution. This means that in case where large amount of data is available the expectation maximization algorithm can reconstruct the distribution with little or almost no information loss.

Very little work has been done on privacy preserving data mining using soft computing methodologies. In past, various heuristic based techniques, cryptography based techniques and reconstruction based techniques have been developed on privacy preserving data mining [32]. In soft computing, study has been done on privacy classification technique on distributed databases including decision tree [34, 35], kNN [36, 37], SVM [38, 39], Bayesian classifier [40] and boosting [41]. Privacy preserving perceptron neural network learning algorithm works in real world on distributed databases [42].

Study Area

The proposed area of work is Soft Computing Techniques based Privacy Preserving Data Mining. We intend to develop/use various Soft Computing techniques like Neural Networks, Fuzzy Logic, Genetic Algorithms, Rough Sets or hybridization of any of these based privacy preserving data mining algorithms. We also intend to perform a comparative study to evaluate the effectiveness of algorithms in privacy preserving data mining.

II. Justification of the proposed work

Soft computing differs from conventional (hard) computing in that, unlike hard computing, it is tolerant of imprecision, uncertainty and partial truth. In effect, the role model for soft computing is the human mind. The guiding principle of soft computing is:

Exploit the tolerance for imprecision, uncertainty and partial truth to achieve tractability, robustness and low solution cost. The employment of soft computing techniques leads to systems which have high MIQ (Machine Intelligence Quotient). In large measure, it is the high MIQ of SC-based systems that accounts for the rapid growth in the number and variety of applications.

To date, a good number of studies have been performed on privacy preserving data mining. Also, there exist a good number of soft computing based data mining algorithms. However, to the best of our knowledge there is very less work done on privacy preserving data mining using soft computing techniques. Moreover, no privacy preserving algorithm exists that outperforms all others on all possible criteria like performance, data utility level of uncertainty or resistance to different data mining algorithms. Rather, an algorithm may perform better than another on one specific criterion [32]. So, it is interesting to study the effectiveness of soft computing based algorithms in privacy preserving data mining.

III. Aims & Objectives:

As discussed earlier, the true problem of privacy intrusion is not with data mining, but with the way data mining is done. The aim of the work is to develop soft computing techniques based privacy preserving data mining algorithms that will enjoy the benefits of data mining but not at the cost of the losing the privacy of the individuals. Particularly Soft Computing techniques have been the thrust area for PPDM because the computational power of the Computers along with high MIQ of the brain (basis of Soft Computing) shall yield intelligent results at a faster rate with high accuracy.

IV. Material proposed to be used for the investigation

Datasets from various relevant organizations.

Paid datasets for research purposes through funding of research projects.

Freely distributable datasets.

Online Digital Libraries for research papers.

GA Libraries

Software packages like MATLAB, WEKA, NN Tool Kit, Fuzzy Tool Box.

V. Methodology proposed to be followed

As the figure 3 shows, privacy-preserving data mining usually has multiple steps that translate to a three-tiered architecture: At the bottom tier are the *data providers*, the data owners, which are often physically distributed. The data providers submit their private data to the *data warehouse server*. This server, which constitutes the middle tier, supports

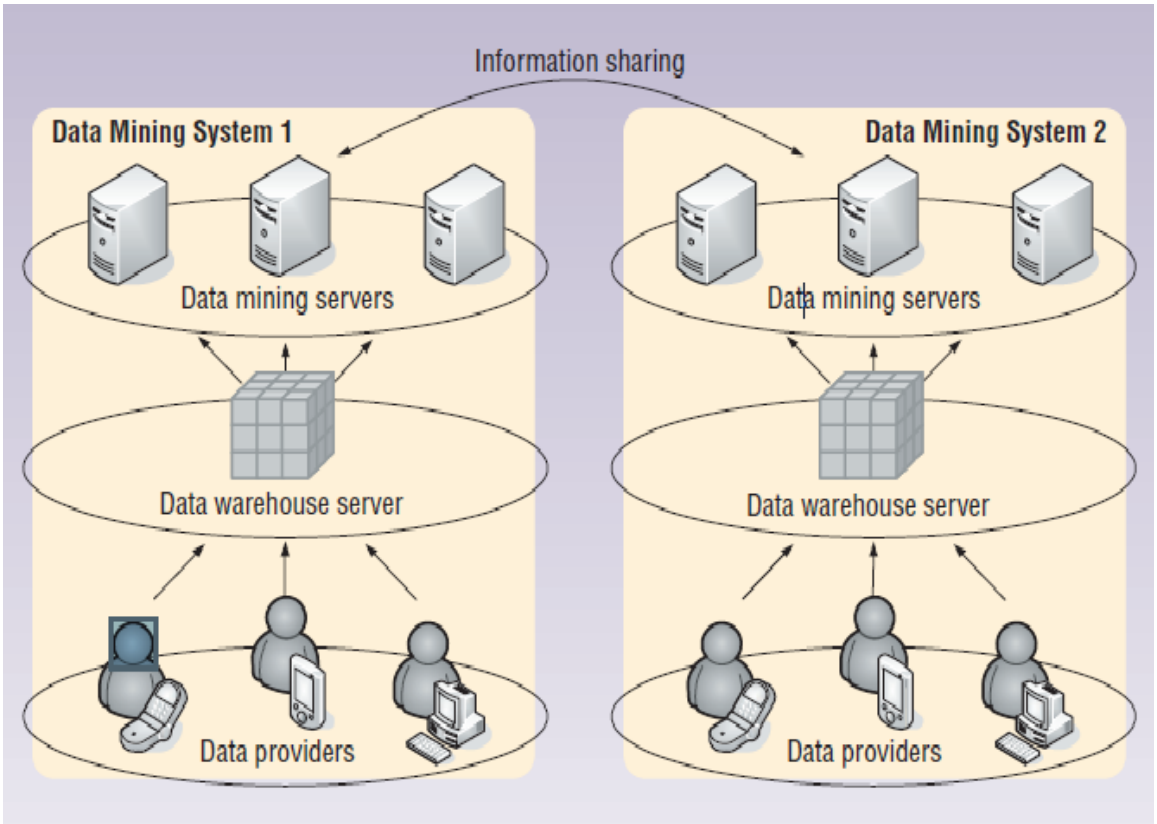

Figure 3: Basic architecture of PPDM

online analytical data processing to facilitate data mining by translating raw data from the data providers into aggregate data that the data mining servers can more quickly process. The data warehouse server stores the data collected in disciplined physical structures, such as a multidimensional data cube, and aggregates and pre-computes the data in various forms, such as sum, average, max, and min. In an online survey system, for example, the survey respondents would be data providers who submit their data to the survey analyzer's data warehouse server; an aggregated data point might be the average age of all survey respondents. The aggregated data is more efficient to process than raw data from the providers. At the top tier are the *data mining servers*, which perform the

actual data mining. In a privacy-preserving data mining system, these servers do not have free access to all data in the data warehouse. In a hospital system, the accounting department can mine patients' financial data, for example, but cannot access patients' medical records. Developing and validating effective rules for the data mining servers' access to the data warehouse is an open research problem. Besides constructing data mining models on its local data warehouse server, a data mining server might share information with data mining servers from other systems. The motivation for this sharing is to build data mining models that span systems. For example, several retail companies might opt to share their local data mining models on customer records to build a global data mining model about consumer behavior that would benefit all the companies. As Figure shows, sharing occurs in the top tier, where each data mining server holds the data mining model of its own system. Thus, "sharing" means sharing local data mining models rather than raw data.

In this work, we intend to study soft computing techniques based privacy preserving data mining. For this, first of all we need to understand the key steps in privacy preserving data mining. Then, we need to identify the steps where we can apply different soft computing techniques, fuzzy logic, neural networks, genetic algorithms, and rough sets. On the basis of this identification, we will apply suitable soft computing techniques for privacy preserving data mining on different data sets of different utility areas for generalizations. Emphasis shall be on quantifying privacy gains and information loss. We shall also be able to provide metrics for testing the effectiveness of PPDM algorithm. Further, we intend to develop the novel algorithms for privacy preserving data mining, which will be based on different soft computing techniques and that shall be able to perform well for most of the criteria like performance, data utility level of uncertainty or resistance to different data mining algorithms. After this, we intend to evaluate the effectiveness of these algorithms on some suitable datasets. Then, we will perform a comparative study of these novel algorithms to identify the most effective algorithms for privacy preserving data mining in different scenarios.

Action Plan and Time Line:

The emphasis of this proposed work shall be on using various Soft computing techniques on different type of datasets, study their behavior as far as the utility of the data mining and the privacy concerns are concerned. It is proposed to carry out the research work over a period of three years and the tentative work plan for each year is as follows:

1. *First Year:*

   During the first year the main task would be to study and understand various soft computing techniques proposed so far in the proposed field of data mining in general and PPDM in particular. Different models and methodologies shall be studied for their advantages and limitations. Efforts will be made to implement some existing soft computing techniques to understand them in some other areas (other than the proposed area of original author).

   In addition all efforts will be made to have a comprehensive literature collection on different aspects of soft computing based PPDM.

2. *Second Year:*

   After the study of various soft computing based algorithms and models for PPDM a set of parameters shall be proposed for the evaluation of algorithms taking into consideration most of the situations. Comparative analysis of different algorithms in the light of the predefined parameters shall be made.

3. *Third Year:*

   Model for PPDM may be proposed or even a novel soft computing based algorithm may be proposed for PPDM.

   Final Reporting of facts and details of background will be presented in the form of book as thesis for completion of research work. Around six months of the third year will be devoted for thesis writing and checkup.

VI. Possible Outcome of the investigation

The possible outcome of the work would be the novel algorithms for privacy preserving data mining, which will ensure that the utility of data mining is not sacrificed and at the same time privacy of secure data is preserved for all possible known criteria. These

algorithms will be based on soft computing techniques and will highlight the benefits of soft computing in privacy preserving data mining.

References

1. Ann Cavoukian, Information and Privacy Commissioner, Ontario, (1997), "***Data Mining Staking a Claim on Your Privacy***", Website: www.ipc.on.ca
2. U. M. Fayyad, G. Piatesky-Shapiro, P. Smyth, and R. Uthurusamy. "***Advances in Knowledge Discovery and Data Mining***". AAAI/MIT Press, 1996.
3. J. Han, M. Kamber. **"*Data Mining: Concepts and Techniques",*** Morgan Kaufmann Publishers, August 2000.
4. Knuggets Software for Data Mining, "***Analytics and Knowledge Discovery",*** http://www.knuggets.com/software/index.html
5. Aris Gkoulalas-Divanis and Vassilios S. Verikios,(2010), "***An Overview of Privacy Preserving Data Mining***", Published by The ACM Student Magazine.
6. The Economist. "***The End of Privacy***", May 1st, 1999. pp: 15
7. C. Clifton and D. Marks. "***Security and Privacy Implications of Data Mining***", ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, 1996. pp: 15-19
8. K. Thearling, "***Data Mining and Privacy: A Conflict in Making***", DS, November 1998.
9. R. Agrawal and R. Srikant. "***Privacy Preserving Data Mining***", ACM SIGMOD Conference on Management of Data, pp: 439-450, 2000
10. Y. Lindell and B. Pinkas, "***Privacy Preserving Data Mining***", Journal of Cryptology, 15(3), pp.36-54, 2000.
11. C. Clifton, M. Kantarcioglu, and J. Vaidya. "***Defining Privacy for Data Mining***", Proceedings of the National Science Foundation Workshop on Next Generation Data Mining, 2002. pp. 126-133.
12. S. R. M. Oliveira and Osmar R. Zaiane, "***Toward Standardization in Privacy-Preserving Data Mining***", DMSSP 2004 (In conjunction with SIGKDD 2004).
13. Sushmita Mitra, Sankar K. Pal and Pabitra Mitra, "***Data Mining in Soft Computing Framework: A Survey***", IEEE Transactions on Neural Networks, Vol 13, No. 1, 2002.
14. L. A. Zadeh, "***Fuzzy Logic, Neural Networks, and Soft Computing.***" *Communications of the ACM*, vol. 37, no. 3, pp: 77-84, March 1994.

15. N. R. Adam and J. C. Wortmann, "***Security Control Methods for Statistical Databases: A Comparison Study***", ACM Comput. Surv. 21(4), 1989. pp: 515-556.
16. D. Agrawal and C. Aggarwal, "***On the Design and Quantification of Privacy Preserving Data Mining Algorithms***", PODS 2001. pp: 247-255.
17. A. Evfimievski, R. Srikant, R. Agrawal and J. Gehrke, "***Privacy Preserving Mining of Association Rules***", SIGKDD 2002. pp. 217-228
18. S. Rizvi and J. Haritsa, "***Maintaining Data Privacy in Association Rule Mining***", VLDB 2002. pp: 682-693.
19. W. Du and Z. Zhan, "***Using Randomized Response Techniques for Privacy Preserving Data Mining****",* SIGKDD 2003. pp. 505-510.
20. Philip Chan, "***An Extensible Meta Learning Approach for scalable and Accurate Inductive Learning***", PhD Thesis, Department of Computer Sciences, Columbia University, New York, NY, 1996
21. Philip Chan, "***On the accuracy of meta-learning for scalable data mining***". Journal of intelligent Information Systems, 8:5-28, 1997.
22. Rong Chen, Krishnamoorthy Sivakumar, and Hillol Kargupta, "***Distributed web mining using bayesian networks from multiple data streams***". In The 2001 IEEE International Conference on Data Mining, San Jose, CA, November 29 - December 2 2001. IEEE.
23. David Wai-Lok Cheung, Vincent Ng, Ada Wai-Chee Fu, and Yongjian Fu, "***Efficient mining of association rules in distributed databases***". IEEE Transaction on Knowledge and Data Engineering, 8(6):911-922, December 1996.
24. Hillol Kargupta and Philip Chan, editors, "***Advances in Distributed and Parallel Knowledge Discovery"***. AAAI/MIT Press, September 2000.
25. Andreas Prodromidis, Philip Chan, and Salvatore Stolfo, "***Advances in Distributed and Parallel Knowledge Discovery***", Chapter 3: Meta-learning in distributed data mining systems: Issues and approaches. AAAI/MIT Press, September 2000.
26. Ashoka Savasere, Edward Omiecinski, and Shamkant B. Navathe, "***An efficient algorithm for mining association rules in large databases".*** In Proceedings of

Twenty first International Conference on Very Large Data Bases, pages 432-444, Zurich, Switzerland, September 11-15 1995. VLDB.

27. Pradeep Shenoy, Jayant R. Haritsa, S. Sundarshan, Gaurav Bhalotia, Mayank Bawa, and Devavrat Shah, "***Turbo-charging vertical mining of large databases"***. In Proceedings of the Nineteenth ACM SIGMOD International Conference on Management of Data, pages 22-33, Dallas, TX, 2000.
28. Rudiger Wirth, Michael Borth, and Jochen Hipp, "***When distribution is part of the semantics: A new problem class for distributed knowledge discovery"***. In Ubiquitous Data Mining for Mobile and Distributed Environments workshop associated with ECML'01 and PKDD'01, Freiburg, Germany, September 3-7 2001.
29. Mohammed J. Zaki, "***Parallel and distributed association mining: A survey***". IEEE Concurrency, special issue on Parallel Mechanisms for Data Mining, 7(4):14{25, December 1999.
30. S. Agrawal and J. Haritsa, "***On Addressing Efficiency Concerns in Privacy-Preserving Mining*"**, DASFAA 2004. pp. 113-124.
31. Stanley, R.M.O. and R.Z Osmar, "***Towards Standardization in Privacy Preserving Data Mining***", Published in Proceedings of 3rd Workshop on Data Mining Standards, WDMS' 2004, USA, p.7-17.
32. Vassilios, S.V., B. Elosa, N.F. Igor, P.P. Loredana, S. Yucel and T. Yannis, 2004, "***State of the Art in Privacy Preserving Data Mining***" Published in SIGMOD Record, 33, 2004, pp: 50-57.
33. Elisa, B., N.F. Igor and P.P. Loredana. "***A Framework for Evaluating Privacy Preserving Data Mining Algorithms***", Published by Data Mining Knowledge Discovery, 2005, pp.121-154.
34. Emekci, O. F., D Sahin, D. Agrawal and A. El Abbadi, "***Privacy Preserving Decision Tree Learning over Multiparties***", Data Knowledge Eng., 2007, 63: pp: 348-361
35. Justin, Z., 2007, "***Using Homomorphic Encryption for Preserving Collaborative Decision Tree Classification***", Proceedings of IEEE Symposium on Computational Intelligence and Data Mining, 2007, pp: 637-645.

36. Mark S., K Yongdae and K. Vipin, "***Privacy Preserving Nearest Neighbour Search***", Proceedings of 6th IEEE International Conference on Data Mining-Workshops, 2006, p: 15-45
37. Artak, A and E. C. Vladimir, "***The Privacy of k-NN Retrieval for Horizontal Partitioned Data: New Methods and Applications***", Proceedings of 18th Conference on Australian Database, Australian Computer Society, 2007, pp: 33-42.
38. Sven L., L. Helger and M Taneli, "***Cryptographically Private Support Vector Machines***", Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, 2006 pp: 618-624
39. Jaideep C., Y Hwanjo and J. Xiaoqian, "***Privacy Preserving SVM classification***", Knowledge Information Systems,2008, 14:161-178
40. Zhiqiang Y, N.H. Rebecca, "***Privacy Preserving Computation of Bayesian Networks on Vertically Partitioned Data***", IEEE Trans. Knowledge Data Engineering, 2006, 18:1253-1264
41. Sebastien G., K. Balazs and A. Esma, "***Privacy Preserving Boosting***", Data Mining Knowledge Discovery, 2007, 14: 131-170.
42. Li Guang, Wang Ya Dang and Su Xiao Hong, "***A Privacy Preserving Neural Network Learning Algorithm for Horizontally Partitioned Databases***", Asian Network for Scientific Information, Informational Technology Journal 9(1): pp: 1-10, 2010.
43. S. Pal, V. Talwar, and P. Mitra, "***Web Mining in Soft Computing Framework: Relevance, State of the Art and Future Directions.***", IEEE Transactions on Neural Networks 13 (5) (2002), pp: 1163–1177.
44. D. Agrawal and C. Agrawal, ***"On design and quantification of Privacy preserving data mining algorithms"***, ACM Magazine, 247-255, 2001.